

Anna Staroń, Anna Grabowska, Elżbieta Katarzyna Jagusztyn-Krynicka*

Zakład Genetyki Bakterii, Instytut Mikrobiologii Uniwersytetu Warszawskiego
02-096 Warszawa, ul. Miecznikowa 1

Wpłynęło w marcu 2008 r.

1. Składniki systemu eksperymentalnego. 1.1. Gospodarz. 1.2. Wektor. 1.2.1. Promotor. 1.2.2. Miejsce wiązania rybosomu. 1.2.3. Znaczniki. 1.2.4. Terminatory. 1.2.5. Markery. 1.2.6. Miejsce startu replikacji. 1.3. Gen. 1.3.1. Używalność kodonów. 1.3.2. Stabilność mRNA. 1.3.3. Reguła N-końca. 2. Strategie eksperymentalne. 2.1. Lokalizacja białka. 2.1.1. Cytoplazma. 2.1.2. Błony komórkowe. 2.1.3. Peryplazma. 2.1.4. Sekrecja do podłoża. 2.2. Strategie klonowania. 2.2.1. Projektowanie konstruktów. 3. Podsumowanie

Overproduction and purification of the recombinant, heterologous proteins from *Escherichia coli* cells

Abstract: Production of sufficient amounts of chemically and conformationally homogenous proteins is a major requirement for basic research (functional and structural studies of proteins) as well as for application studies (immunoprophylaxis and therapy). Over the past twenty years, numerous expression systems (based on prokaryotic, yeast, insect, plant and mammalian cell cultures) have been described and tested for protein overproduction and purification. Here, we describe the major recent advances relevant to the successful overproduction and purification of heterologous proteins in the *Escherichia coli* system which is, thus far, the most commonly used organism for heterologous protein production. Issues addressed in this review include: vectors used for recombinant protein expression, cloning strategies and using different fusion tags that enhance protein solubility and facilitate protein purification. The influence of recombinant protein localization on enhancing the total protein yield and on retaining their native conformation is also discussed.

1. Elements of the experimental system. 1.1. Host. 1.2. Cloning vector. 1.2.1. Promoter. 1.2.2. Ribosom binding site. 1.2.3. Tags. 1.2.4. Terminators. 1.2.5. Selective markers. 1.2.6. Origin of replication. 1.3. Gene. 1.3.1. Codon usage. 1.3.2. RNA stability. 1.3.3. N-end rule. 2. Experimental strategies. 2.1 Protein localization 2.1.1. Cytoplasm. 2.1.2. Cell membranes. 2.1.3. Periplasm. 2.1.4. Secretion to the medium. 2.2. Cloning strategies. 2.2.1. Construct design. 3. Summary

Słowa kluczowe: ekspresja, heterologiczne białka, nadprodukcja, oczyszczanie

Key words: expression, heterologous proteins, overproduction, purification

1. Składniki systemu eksperymentalnego

Zaledwie niewielki procent białek organizmów żywych występuje w komórkach w ilości wystarczającej, by umożliwić ich bezpośrednią izolację na dużą skalę. Aby otrzymać znaczące ilości białka do celów poznawczych (analizy funkcjonalne i strukturalne) lub aplikacyjnych (terapia i profilaktyka chorób organicznych i zakaźnych), zazwyczaj konieczna jest jego nadprodukcja, do której osiągnięcia niezbędne są trzy elementy – dobrany gospodarz, właściwy wektor i odpowiednio namnożony gen/fragment genu, kodujący oczyszczane białko.

1.1. Gospodarz

Do otrzymywania rekombinowanych białek najczęściej stosowane są systemy bakteryjne, w szczególności oparte na hodowli *Escherichia coli*, modelowego orga-

nizmu stosowanego w eksperymentach inżynierii genetycznej. W tym przypadku zaletą *E. coli* jest zdolność szybkiego wzrostu na tanich substratach, wysoki stopień scharakteryzowania genomu oraz dostępność szerokiej gamy wektorów i szczepów. *E. coli* jest gatunkiem niepatogennym [6], hodowanym w prawie każdym laboratorium, stąd metody hodowli, także na dużą skalę, są dobrze opracowane.

W laboratoriach rutynowo stosuje się szczepy K12 lub B (np. BL21/DE3 (*lon*, *ompT*) oraz pochodzące od nich mutanty (Tab. I) [41].

Produkcja rekombinowanych białek w komórkach *E. coli* ma jednak kilka poważnych ograniczeń. U tego gatunku bakterii nie funkcjonują niektóre mechanizmy modyfikacji posttranslacyjnych (np. glikozylacja) syntetyzowanych białek, podczas gdy modyfikacje te mogą być kluczowe dla prawidłowej aktywności białka. [22]. Część problemów dotyczących wprowadzania modyfikacji posttranslacyjnych w *E. coli* może nie-
długo doczekać się rozwiązania. W ostatnich latach

* Autor korespondencyjny: Zakład Genetyki Bakterii, Instytut Mikrobiologii Uniwersytetu Warszawskiego, 02-096 Warszawa, ul. Miecznikowa 1; tel. (22) 554 12 16; e-mail: kjkryn@biol.uw.edu.pl

Tabela I

Szczepy *E. coli* najczęściej stosowane do produkcji rekombinowanych białek

Szczep <i>E. coli</i>	Pochodzenie	Główne cechy
BL21 <i>trxB</i>	BL21	<i>trxB</i> – szczep umożliwiający wprowadzanie mostków disiarczkowych w cytoplazmie
	BL21	szczep posiadający niescharakteryzowaną mutację umożliwiającą nadekspresję białek błonowych
JM 83	K-12	stosowany do nadekspresji białek o lokalizacji peryplazmatycznej
Origami/Origami B	K-12/BL21	<i>trxB/gor</i> – szczep umożliwiający wprowadzanie mostków disiarczkowych w cytoplazmie
Rosetta	BL21	szczep umożliwiający nadprodukcję białek kodowanych przez geny zawierające kodony rzadko używane w <i>E. coli</i> ^a
Rosetta-gami	BL21	<i>trxB/gor</i> – szczep umożliwiający wprowadzanie mostków disiarczkowych w cytoplazmie, ułatwia ekspresję białek kodowanych przez geny zawierające kodony rzadko używane w <i>E. coli</i> ^a

^a AUA, AGG, AGA, CGG, CUA, CCC oraz GGA

w *E. coli* sklonowano zespół siedemnastu genów *pgl* (ang. protein glycosylation) z komórek *Campylobacter jejuni*, odpowiedzialnych za proces N-glikozylacji [44]. System ten okazał się funkcjonalny w komórkach biocy. Szczegółowa charakterystyka tego szlaku enzymatycznego może w przyszłości umożliwić otrzymywanie glikozylowanych białek także w komórkach *E. coli*.

Niektórych białek, szczególnie błonowych, nie udało się otrzymać w *E. coli*, albo też ich nadprodukcja prowadzi do ich akumulacji w postaci nierozpuszczalnej – tzw. ciał inkluzyjnych. W przypadku terapeutycznego stosowania oczyszczonego białka problemem może być obecność w preparacie lipopolisacharydu (LPS, składnika błony zewnętrznej bakterii Gramujemnych), który ma działanie pirogenne. Wytwarzane w *E. coli* białka o masie cząsteczkowej większej niż 30 kDa mogą być nieprawidłowo fałdowane [17]. Niepatogenne szczepy *E. coli* nie posiadają także efektywnych mechanizmów sekrecji białek do podłoża.

Alternatywnymi dla *E. coli* gospodarzami produkującymi heterologiczne białka są inne gatunki bakterii, m.in. gramodatnie mikroorganizmy z rodzaju *Bacillus*, które w odróżnieniu od *E. coli* nie syntetyzują błony zewnętrznej, a to umożliwia późniejsze terapeutyczne zastosowanie preparatów. Rodzaj *Bacillus* jest zdolny do sekrecji dużych ilości białek do podłoża, co znacznie ułatwia ich oczyszczanie. W odróżnieniu od *E. coli* niewiele wiadomo o tworzeniu mostków disiarczkowych u bakterii tego rodzaju [41]. Inne stosowane w eksperymentach nadekspresji gatunki to m.in. *Lactococcus lactis*, stosowany z sukcesem do nadprodukcji białek błonowych [21, 22], *Corynebacterium glutamicum* [35], czy też bakterie z rodzaju *Streptomyces* [41].

Żaden z wyżej wspomnianych gatunków bakterii nie jest w stanie wprowadzać pewnych modyfikacji posttranslacyjnych do produkowanych białek. Kiedy modyfikacje posttranslacyjne są istotne dla aktywności białka, stosuje się systemy inne niż bakteryjne. Mogą być

to grzyby z typu workowców, w tym drożdże (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* czy *Pichia pastoris*) lub grzyby strzępkowe (*Aspergillus niger*, *A. oryzae*, *Trichoderma reesei*). Posiadają one wiele wspólnych zalet z systemami bakteryjnymi, jak np. zdolność do szybkiego wzrostu i stosunkowo niski koszt hodowli. Wydzielanie nadekspresjonowanych białek do podłoża jest w niektórych systemach znacznie ułatwione [22, 26]. Jednakże, szczególnie w przypadku białek ludzkich produkowanych w komórkach grzybów, otrzymane produkty mogą mieć niepożądane właściwości (np. być immunogenne) czego przyczyną są różnice w procesowaniu białek w grzybach w stosunku do organizmów wyższych [12].

Do nadprodukcji białek wykorzystywane są także komórki owadzie (system bakulowirusowy), szczególnie efektywne w sekrecji białek do podłoża [17], oraz różnorodne, głównie mysie lub ludzkie linie komórek ssaczy. W obu tych systemach do syntetyzowanych białek wprowadzane są modyfikacje posttranslacyjne, co zwiększa szanse na otrzymanie końcowego produktu o prawidłowej strukturze i wykazującego wysoką aktywność. Komórki ssaków wydają się być lepszym gospodarzem do otrzymywania białek błonowych niż *E. coli*, między innymi ze względu na wolniejsze tempo translacji i fałdowania w porównaniu z komórkami mikroorganizmów [47]. Wadą tej strategii jest jednakże dużo niższa wydajność i wyższe koszty.

Wykorzystanie zwierząt transgenicznych otwiera możliwości produkcji białek zawierających odpowiednie modyfikacje posttranslacyjne na dużą skalę. Białka produkuje się w mleku, krwi, czy nasieniu. Jednak bardzo wysokie koszty, niska wydajność oraz kontrowersje bioetyczne, jakie budzą takie metody otrzymywania białek sprawiają, że nie jest to system stosowany powszechnie i na dużą skalę [12].

W dalszej części pracy przeglądowej będą omówione jedynie zagadnienia związane z wytwarzaniem białek w komórkach *E. coli*.

1.2. Wektor

1.2.1. Promotor

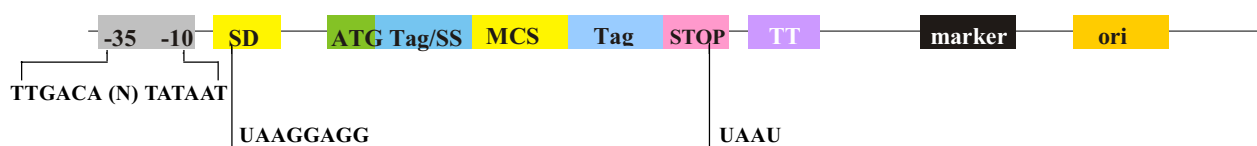
Opisano bardzo wiele promotorów, które mogą być potencjalnie używane do nadprodukcji białek w *E. coli*, jednak tylko niektóre z nich są stosowane powszechnie [41]. Promotor do konstrukcji wektora ekspresyjnego powinien być silny (dla większości zastosowań warto, aby produkowane białko stanowiło 10–30% wszystkich białek w komórce; [16]). Ważne jest zminimalizowanie ekspresji podstawowej (przy braku induktora) szczególnie, gdy istnieje podejrzenie toksyczności syntetyzowanego białka dla gospodarza. Brak skutecznej represji może powodować niestabilność plazmidu, zmniejszenie tempa wzrostu hodowli i generowanie mutacji w ekspresowanym genie [23]. Promotor musi być indukowalny za pomocą prostych i tanich substratów, niestosowanych w standardowych podłożach [23].

Wiele systemów ekspresyjnych zostało skonstruowanych w oparciu o dobrze znany mechanizm regulacji operonu laktozowego *E. coli*. Promotor *lacZp* jest stosunkowo słabym promotorem i w związku z tym jest rzadko wykorzystywany do produkcji rekombinowanych białek. Ta jego cecha może być zaletą w przypadku nadprodukcji białek błonowych lub innych białek toksycznych dla gospodarza. Promotor *lacUV5p* stanowi nieco silniejszy wariant *lacZp*, charakteryzujący się zmniejszoną czułością na cAMP, a przez to skuteczniejszą kontrolą ekspresji podstawowej [3]. Promotory *tacp* i *trcp* to dwa syntetyczne warianty silniejsze od *lacZp*, różniące się między sobą jednym nukleotydem. Składają się z regionu -35 promotora *trp* i regionu -10 promotora *lacZp* [3].

Ekspresja z promotora *lacZp* i jego pochodnych jest regulowana przez represor LacI. W warunkach nieobecności lub bardzo niskich stężeń induktora (allo-laktozy bądź IPTG) represor wiąże się do sekwencji nukleotydowej operatora zachodzącej na promotor i uniemożliwia transkrypcję. Jednak poziom represora LacI w typie dzikim nie jest wystarczający do całkowitej represji przy braku IPTG. Z tego powodu skonstruowano dwa szczepy ze zmienioną sekwencją nukleotydową -35 promotora *lacIp*, nazwane *lacI^q* oraz *lacI^l*, produkujące większą ilość represora LacI [32].

Jednym z najczęściej używanych jest system polimerazy RNA faga T7 [38]. Polimeraza RNA faga T7 transkrybuje około 230 nukleotydów na sekundę, jest więc pięć razy bardziej efektywna niż polimeraza RNA *E. coli*, która jest w stanie w tym samym czasie przeprowadzić transkrypcję około 50 nukleotydów. Te dwie polimerazy RNA rozpoznają zupełnie odmienne sekwencje promotorowe [34]. Ekspresyjny gen klonuje się pod kontrolą silnego promotora bakteriofaga T7 na plazmidzie o średniej liczbie kopii (np. wektory serii pET). Polimeraza RNA faga T7 jest dostarczana *in trans*. Gospodarz posiada wbudowany do swojego chromosomu genom faga λ, zawierający gen kodujący polimerazę RNA faga T7 pod kontrolą promotora *lacUV5p* (λDE3) [3]. Ze względu na bardzo niski poziom ekspresji genu kodującego polimerazę RNA faga T7 przy braku induktora, kontrola ekspresji jest ścisła. Często w rejonie promotorowym wprowadza się dodatkowo sekwencję nukleotydową operatora *lacOo* wiążącą represor laktozowy LacI (promotor *T7lacp*). Istnieje możliwość jeszcze silniejszego obniżenia ekspresji podstawowej poprzez zastosowanie wektorów niosących geny kodujące lizozym faga T7, będący inhibitorem polimerazy RNA faga T7 (wektory pLysS czy pLysE). Szczepy niosące te plazmidy mają jednak obniżone tempo wzrostu [41].

Systemy wykorzystujące opisane powyżej promotory mogą być indukowane za pomocą IPTG. Jest to jednak metoda kosztowna, szczególnie w przypadku produkcji na dużą skalę, a także, ze względu na toksyczność IPTG, niekorzystna w przypadku późniejszych zastosowań terapeutycznych oczyszczonego białka [23]. Alternatywą dla indukcji IPTG jest opracowana stosunkowo niedawno metoda autoindukcji [39], oparta na zdolności bakterii do wykorzystywania różnych źródeł węgla i do negatywnej regulacji ekspresji genów na drodze represji katabolicznej. W podłożu wzrostowym obecne SA dwa cukry: glukoza (0,05%) i laktoza (0,2%). Na początku wzrostu hodowli glukoza stanowi źródło węgla i zarazem represor genów pozostających pod kontrolą promotora *lacZp* bądź jego pochodnych, uniemożliwia także pobieranie laktozy do komórki przez permeazę laktozową. Gdy hodowla osiągnie fazę stacjonarną, poziom



Rys. 1. Schemat budowy bakteryjnego wektora ekspresyjnego.

Jako przykład pokazano sekwencje nukleotydowe -35 i -10 promotora *tac*. SD – sekwencja Shine-Dalgarno. ATG – najczęściej stosowany w wektorach ekspresyjnych kodon startu translacji. Tag – znacznik dołączony do ekspresowanego białka; SS – sekwencja sygnałowa; MCS – polilinker (multiple cloning site); TT – terminator transkrypcji; marker – najczęściej gen warunkujący oporność na antybiotyki; ori – miejsce startu replikacji DNA plazmidowego (origin of replication).

glukozy w podłożu ulega obniżeniu. Represja kataboliczna zostaje zniesiona, powodując indukcję promotora *lacZp*; laktoza staje się głównym źródłem węgla. Jedną z zalet tej metody jest wzrost hodowli do osiągnięcia dużej gęstości przed zaindukowaniem ekspresji. Dzięki temu możliwe jest uzyskanie dużej ilości białek, także tych toksycznych dla *E. coli* [28].

Zastosowanie promotora bakteriofaga T7 pozwala także na zmniejszenie ilości białek komórkowych poprzez selektywne zahamowanie aktywności polimerazy RNA *E. coli* za pomocą rifampicyny dodawanej do hodowli jednocześnie z induktorem [14]. Ekspresja genu transkrybowanego przez polimerazę RNA bakteriofaga T7 nie zostaje w ten sposób zahamowana.

Innym stosowanym często promotorem jest promotor P_{BAD} . W tym przypadku ekspresja genów jest kontrolowana przez aktywator AraC i indukowana po dodaniu do podłoża arabinozy. Zalety tego systemu to niskie koszty indukcji, wysoki poziom ekspresji genu, skuteczna regulacja i możliwość stopniowania poziomu ekspresji, co odróżnia tę strategię od pozostałych systemów ekspresyjnych [41]. Jednak w tym przypadku wydajność jest dużo niższa niż systemu wykorzystującego polimerazę RNA faga T7 [28].

Tabela II podaje sekwencje nukleotydowe promotorów obecnych w najczęściej stosowanych wektorach ekspresyjnych, rozpoznawane przez główną podjednostkę σ^{70} RNAP.

Tabela II
Sekwencje nukleotydowe najczęściej stosowanych w wektorach ekspresyjnych promotorów rozpoznawanych przez główną podjednostkę σ^{70} RNAP. Nukleotydy zgodne z sekwencją (consensus) dla σ^{70} pogrubiono

Promotor	Region -35	Odstęp	Region -10
<i>lacZp</i>	TTtACA	18	TATgaT
<i>lacUV5p</i>	TTtACA	18	TATAAT
<i>trpp</i>	TTGACA	17	TtaAcT
<i>tacp</i>	TTGACA	17	TATAAT
P_{BAD}	cTGACg	17	TActgT
δ^{70} konsensus	TTGACA	17	TATAAT

1.2.2. Miejsce wiązania rybosomu

W komórkach bakteryjnych inicjacja translacji wymaga obecności w mRNA miejsca wiązania rybosomu (RBS), zawierającego sekwencję Shine-Dalgarno (SD) i kodon start translacji. Miejsce wiązania rybosomu (RBS) obejmuje odcinek ok. 54 nukleotydów pomiędzy fragmentem -35 promotora i nukleotydem +20 (± 2) [23]. Fragment Shine-Dalgarno (SD), którego optymalną sekwencją nukleotydową jest 5'-UAAGGAGG-3', oddziałuje z 3' końcem 16S rRNA podczas inicjacji translacji. Umiejscowiony jest 7 ± 2

nukleotydy powyżej kodonu startu translacji, którym w większości systemów ekspresyjnych jest kodon AUG [34]. Znaczący wpływ na wydajność translacji ma też kodon następujący po kodonie start [36].

1.2.3. Znaczniki

Dla ułatwienia oczyszczania, detekcji czy też zwiększenia rozpuszczalności rekombinowanych białek dołączany jest do nich często znacznik – może być nim krótki peptyd, domena bądź całe białko. Znacznik taki powinien spełniać kilka kryteriów, w tym wysokie powinowactwo do złoża stosowanego w procesie oczyszczania, możliwość elucji substancjami niskocząsteczkowymi, niewielki wpływ na aktywność i strukturę trzeciorzędową białka, oraz łatwą wykrywalność. Często w celu otrzymania natywnych białek znaczniki są odcinane podczas oczyszczania białka za pomocą działającej specyficznie proteazy [42].

Dostępna jest cała gama znaczników, zarówno białek łączących się do niskocząsteczkowych ligandów, jak i krótkich peptydów rozpoznawanych przez unieruchomione na złożu białka. Każdy z nich ma swoje wady i zalety. Nie ma idealnego znacznika, który zwiększałby wydajność produkcji białka, jego rozpuszczalność, a jednocześnie ułatwiał oczyszczanie. Często w celu optymalizacji nadprodukcji białka korzystne jest zastosowanie przynajmniej dwóch znaczników [46].

Znaczniki mogą być dołączane do N- lub C-końca białka. Fuzje N-końcowe mają tę zaletę, że zwiększają wydajność produkcji białka poprzez efektywną inicjację translacji. Dołączanie znacznika zarówno na N-, jak i na C-końcu białka może też zmniejszyć szybkość jego degradacji [46].

Ważną właściwością niektórych znaczników jest zwiększanie rozpuszczalności związanych z nimi białek. Według danych z laboratoriów zajmujących się badaniami strukturalnymi, nieco więcej niż połowa rekombinowanych białek wytwarzanych w *E. coli* tworzy nierozpuszczalne agregaty [7]. Bardziej pesymistyczne źródła szacują procent otrzymanych rozpuszczalnych białek na 13–23% [13]. Jednym z możliwych rozwiązań tego problemu, poza obniżeniem temperatury, w jakiej produkowane jest białko, lub zmianą systemu ekspresyjnego np. na system drożdżowy czy bakulowirusowy, jest dołączenie do białka znacznika zwiększającego jego rozpuszczalność. Najczęściej stosowane w takich wypadkach znaczniki to pochodzące z *E. coli* białko wiążące maltozę (maltose-binding protein, MBP) i „N-utilization substance A” (NusA) [46].

Stopień rozpuszczalności białka zlokalizowanego w cytoplazmie *E. coli* można oszacować stosując model Wilkinsona i Harrisona (dostępny na stronie internetowej <http://www.biotech.ou.edu>). Między innymi dzięki teoretycznym wyliczeniom opartym

na tym modelu zidentyfikowano NusA jako białko o wysokiej rozpuszczalności [34].

Porównanie najczęściej stosowanych znaczników przedstawia Tabela III.

Ciekawym rozwiązaniem są tzw. rainbow tags, znaczniki nadające sygnał wizualny (barwę, świecenie), stosowane do śledzenia losów białka podczas procedury oczyszczania. Stosowana w tym celu domena wiążąca mononukleotydy flawinowe z reduktazy cytochromu P450 (kolor niebiesko-zielony lub żółty) czy też cytochrom b5 (kolor czerwony) są widoczne gołym okiem i nie wymagają do detekcji zewnętrznego źródła energii, w odróżnieniu od stosowanego także GFP (zielone białko fluoryzujące, ang. green fluorescent protein). Znaczniki typu „rainbow” muszą być jednak stosowane w połączeniu z innymi, wykazującymi powinowactwo do złoża [1], a ty samym ułatwiającymi proces oczyszczania.

Pomimo wielu zalet znaczników ich obecność może wpływać na strukturę i aktywność rekombinowanego białka [46]. W wielu przypadkach po oczyszczeniu białka konieczne jest usunięcie znacznika. Proces ten

jest przeprowadzany przez specyficzne endopeptydazy, i stanowi najsłabsze ogniwo metody stosowania znaczników do oczyszczania białek. Wydajność odcinania znacznika może być bardzo różna i powodować zmniejszenie ilości oczyszczonego białka. Optymalizacja warunków tego procesu wymaga dużego nakładu pracy, a otrzymane białko może nie być aktywne lub wytrącać się z roztworu [5]. Cechy najczęściej stosowanych proteaz przedstawiono w tabeli IV.

W przypadku dołączenia znacznika na N-końcu białka możliwe jest uzyskanie natywnego N-końca, jednak tylko z zastosowaniem mniej specyficznych proteaz – czynnika Xa i enterokinazy. Proteazy bardziej specyficzne pozostawiają 1–2 aminokwasy na N-końcu oczyszczonego białka. W przypadku znacznika na C-końcu białka należy liczyć się z pozostawieniem dodatkowych 4–6 aminokwasów na C-końcu przez proteazę, niezależnie od jej specyfiki. Jakkolwiek obecność kilku dodatkowych aminokwasów na ogół nie stanowi problemu, w niektórych przypadkach (np. badania struktury bądź w zastosowaniach terapeutycznych) może dyskwalifikować rekombinowane białko [46].

Tabela III

Porównanie najczęściej stosowanych znaczników. Sporządzono wg [42, 46]

Znacznik	Wielkość (kDa)	Zalety	Wady
His-Tag ^a	0.84	Małe obciążenie metaboliczne Niskie koszty złoża Łagodne warunki elucji Umożliwia oczyszczanie w warunkach natywnych i denaturujących	Specyficzność IMAC ^b jest mniejsza niż innych metod chromatografii powinowactwa Nie zwiększa rozpuszczalności
FLAG-Tag	1.01	Małe obciążenie metaboliczne Wysoka specyficzność	Wysokie koszty złoża Surowe warunki elucji
STREP-Tag ^c	1.05	Małe obciążenie metaboliczne Wysoka specyficzność Łagodne warunki elucji	Wysokie koszty złoża Nie zwiększa rozpuszczalności
S-Tag	1.75	Małe obciążenie metaboliczne Wysoka specyficzność	Wysokie koszty złoża Surowe warunki elucji Nie zwiększa rozpuszczalności
CBP ^d	2.96	Małe obciążenie metaboliczne Wysoka specyficzność Łagodne warunki elucji	Wysokie koszty złoża Nie zwiększa rozpuszczalności
GST-Tag ^e	26.00	Wydajna inicjacja translacji Niskie koszty złoża Łagodne warunki elucji Nie zwiększa rozpuszczalności	Wysokie obciążenie metaboliczne Białko tworzące homodimery
MBP-Tag ^f	40.00	Wydajna inicjacja translacji Niskie koszty złoża Zwiększa rozpuszczalność Łagodne warunki elucji	Wysokie obciążenie metaboliczne
NusA ^g	54.80	Wydajna inicjacja translacji Zwiększa rozpuszczalność	Nie używane do oczyszczania białek, może być stosowane do immunodetekcji

^a His-Tag – znacznik polihistydynowy;

^b IMAC – chromatografia powinowactwa na unieruchomionych jonach metali (ang. immobilized metal affinity chromatography);

^c STREP-Tag – peptyd wiążący streptawidynę; ^d CBP – peptyd wiążący kalmodulinę (ang. calmodulin-binding peptide);

^e GST – S-transferaza glutationu (ang. glutathione S-transferase); ^f MBP – białko wiążące maltozę (ang. maltose-binding protein);

^g NusA – ang. N-utilization substance A

Tabela IV

Porównanie proteaz najczęściej stosowanych do odcinania znaczników w rekombinowanych białkach.
Sporządzono na podstawie [19, 40, 46]

Proteaza	Rozpoznawana sekwencja	Cechy
Enterokinaza	DDDDK/X ^a	Obserwowane niespecyficzne cięcia FLAG-Tag ma wewnętrzne miejsce rozpoznawane przez enterokinazę
Trombina	LVPR/GS ^b	Obserwowane niespecyficzne cięcia Wydajność cięcia można zwiększyć poprzez umieszczenie pięciu reszt glicyny pomiędzy miejscem cięcia a N-końcowym znacznikiem
Czynnik Xa	IE[D]GR/X ^c	Mniej efektywny i mniej specyficzny niż inne peptydazy
Proteaza TEV ^d	EXXYXQ/S ^c	Wysoka specyficzność Wydajne cięcie
PreScission	LEVLFQ/GP	Wysoka specyficzność Wydajne cięcie

^a wydajność cięcia zależy od aminokwasu w pozycji X

^b trombina rozpoznaje także inne sekwencje o strukturze X₁X₂PR[K]X₃X₄, gdzie X₁ i X₂ to aminokwasy hydrofobowe, a X₃ i X₄ nie są kwaśnymi aminokwasami

^c X może być dowolnym aminokwasem za wyjątkiem argininy i proliny

^d pochodząca z wirusa wżerkowej plamistości tytoniu (ang. tobacco etch virus)

^e nie każdy aminokwas jest tolerowany w miejscu X; optymalna sekwencja to ENLYFQS

Rozwiązaniem problemu wydajności odcinania znacznika i otrzymania natywnego N-końca białka może być zastosowanie technologii SUMO. Rodzina białek SUMO jest szeroko rozpowszechniona wśród organizmów eukariotycznych, nie występuje natomiast u bakterii [5]. Białka z tej rodziny są dołączane do innych białek kowalencyjnie, podobnie jak ubikwityna. Dołączenie białka SUMO jest procesem odwracalnym, a za jego odłączenie odpowiedzialne są proteazy SUMO. Białko SUMO w połączeniu ze znacznikiem histydynowym znalazło zastosowanie jako kombinowany znacznik, jednocześnie zwiększający rozpuszczalność białka (białko SUMO) i ułatwiający jego oczyszczanie (znacznik 6xHis). Zaletą tej metody jest możliwość odcięcia znacznika przez proteazę SUMO, która, w odróżnieniu od innych często stosowanych proteaz, rozpoznaje trzeciorzędową strukturę znacznika, a nie konkretną sekwencję aminokwasową. Dzięki temu możliwe jest uzyskanie natywnego N-końca rekombinowanego białka; jedynym wyjątkiem są białka zawierające na N-końcu prolinę. Dodatkowo, proteaza SUMO jest w stanie efektywnie i specyficznie działać w szerokim zakresie temperatury, pH i siły jonowej. Eksperymenty porównujące wydajność działania znaczników przy oczyszczaniu rekombinowanych białek udokumentowały wyższą skuteczność SUMO niż takich znaczników jak MBP, GST czy NusA [24]. Jakkolwiek, jak wspomniano wcześniej, żaden system dołączanych znaczników nie jest idealny, białko SUMO wydaje się być obiecującym rozwiązaniem pozwalającym na przewyciężenie niektórych przeszkód dotychczas napotykanych podczas stosowania znaczników.

1.2.4. Terminatory procesu transkrypcji i translacji

Terminator transkrypcji umieszczony poniżej nukleotydowej sekwencji kodującej nadekspymowane białko zwiększa stabilność plazmidu zapobiegając transkrypcji przez miejsce startu replikacji. Terminacja transkrypcji *E. coli* może być Rho-niezależna i Rho-zależna [32]. Pierwsza klasa terminatorów składa się z odwróconych powtórzeń sekwencji nukleotydowych, po których następuje ciąg nukleotydów adenylowych. Po transkrypcji przez polimerazę RNA sekwencji nukleotydowej odwróconego powtórzenia, mRNA tworzy strukturę szpilki do włosów, co powoduje zatrzymanie się enzymu. Ponieważ zaraz za odwróconym powtórzeniem następuje ciąg nukleotydów urydylowych, które słabo oddziałują z nukleotydami adenyłowymi na nici matrycowej, polimeraza oddysocjuje. W wektorach ekspresyjnych umieszcza się często dwa takie terminatory. Struktura szpilki do włosów na końcu 3' spełnia także funkcję stabilizującą mRNA, chroniąc go przed degradacją i wydłużając jego okres półtrwania [23]. Mechanizmy Rho-zależnej terminacji transkrypcji nie są wykorzystywane w wektorach ekspresyjnych [32].

Terminatory transkrypcji umieszcza się także powyżej ekspymowanego genu, żeby zminimalizować ekspresję podstawową [16].

Terminacja translacji jest najbardziej efektywna w przypadku zastosowania przedłużonego kodonu stop UAAU [29] lub kilku kolejnych kodonów stop, jeden za drugim [34].

1.2.5. Markery selekcyjne

Najczęściej stosowane w wektorach ekspresyjnych markery to geny warunkujące oporność komórek na

ampicylinę, kanamycynę, chloramfenikol i tetracyklinę. Ampicylina jest najmniej praktycznym spośród wymienionych antybiotyków, gdyż łatwo ulega hydrolizie w kwaśnym środowisku, zaś pH hodowli obniża się z czasem wzrostu hodowli bakteryjnej. Gdy eksprymowane białko jest toksyczne, a wektor je kodujący niesie gen warunkujący oporność na ampicylinę, plazmid może nie utrzymywać się stabilnie w populacji. Rozwiązaniem w tym przypadku jest użycie karbenicyliny. W niektórych zastosowaniach terapeutycznych oczyszczonego białka stosowanie antybiotyków może być niepożądane. Alternatywne strategie obejmują m.in. zastosowanie systemów trucizna-antidotum (np. system *hok/sok* z plazmidu R1) w celu zapewnienia stabilnego rozdziału plazmidu [3].

1.2.6. Miejsce startu replikacji

Większość wektorów ekspresyjnych to pochodne plazmidów pBR322 (wektory typu pET, 15–20 kopii na komórkę) lub pUC (część wektorów typu pQE, 500–700 kopii na komórkę). Jeżeli celem jest koekspresja dodatkowych genów, wektory te łączy się w jednym systemie ekspresyjnym ze zgodnymi z nimi pochodnymi plazmidu pACYC184 (pRARE niosący geny kodujące rzadkie tRNA, 10–12 kopii na komórkę) [34].

1.3. Gen

1.3.1. Używalność kodonów

Przy ekspresji genów kodujących rekombinowane białka ważne są nie tylko cechy stosowanego wektora i gospodarza, ale także eksprymowanego genu. Poważnym problemem, który może utrudnić lub nawet uniemożliwić ekspresję, jest różna używalność kodonów w różnych organizmach. Ze względu na fakt, iż 20 aminokwasów kodowanych jest przez aż 61 kodonów, niektórym aminokwasom odpowiada więcej niż jeden kodon (sześć w przypadku Arg, Leu czy Ser) [15]. W każdym organizmie część synonimicznych kodonów jest używana częściej niż pozostałe, a co się z tym wiąże, pula dostępnych tRNA różni się w zależności od organizmu [41]. Z tego względu heterologiczne geny zawierające znaczny procent kodonów rzadko używanych w *E. coli* mogą być eksprymowane z małą

wydajnością [16]. Przykłady różnic w częstoci występowania wybranych kodonów u *E. coli*, *Campylobacter jejuni* i człowieka przedstawiono w tabeli V.

Jednym z bardziej spektakularnych przykładów trudności wynikających z odstępstw od tradycyjnego kodu genetycznego jest produkcja rekombinowanych białek orzęsków (*Ciliata*). System translacji u orzęsków odczytuje kanoniczne kodony stop (TAA i TAG) jako kodony kodujące glutaminian. Ekspresja genów *Ciliata* w *E. coli* bez zmiany nukleotydowej sekwencji kodującej jest z tego powodu niemożliwa [15].

Jedną ze strategii pozwalających na przynajmniej częściowe ominięcie problemu różnej używalności kodonów jest zmiana warunków hodowli, w tym zmiany w składzie podłoża i obniżenie temperatury. Jest to jednak rozwiązanie nie zawsze skuteczne [41].

Najbardziej pracochłonnym rozwiązaniem jest zmiana kodonów w genie kodującym białko na optymalne dla *E. coli*. Można do tego celu zastosować mutagenезę specyficzną co do miejsca, bądź syntezę całego genu *de novo*. Metody takie stosuje się najczęściej przy produkcji białek wirusowych (używalność kodonów u wirusów jest często determinowana przez obecność zachodzących na siebie otwartych ramek odczytu), czy pochodzących z organizmów stosujących odmienny od kanonicznego kod genetyczny, jak wspomniane wcześniej orzęski [15]. Projektowanie nowej sekwencji nukleotydowej genu jest procesem skomplikowanym i wieloetapowym (w przypadku białka o długości 100 aminokwasów istnieje teoretyczna możliwość zastosowania $\approx 5 \times 10^{47}$ sekwencji DNA [15], w którym uwzględnia się nie tylko używalność kodonów, ale także eliminuje nukleotydowe sekwencje powtórzone, sekwencje mogące potencjalnie tworzyć struktury drugorzędowe w mRNA, oraz miejsca restrykcyjne. Do projektowania nukleotydowych sekwencji genów stosuje się specjalne programy, takie jak Gene Designer [43]. Powyższa strategia jest jednak kosztowna i pracochłonna. Najczęściej stosowanym rozwiązaniem jest zastosowanie szczepu *E. coli* o wzbogaconej puli tRNA [34]. Do dostępnych komercyjnie szczepów, takich jak Rosetta czy BL21-CodonPlus wprowadzono plazmidy niosące geny kodujące rzadkie tRNA (rozpoznające kodony AUA, AGG, AGA, CUA, CCC, a w przypadku Rosetta także GGA) [27, 37]. Pomimo pewnych wad (duże obciążenie metaboliczne komórki, niedostateczna modyfikacja nadprodukowanych cząsteczek tRNA prowadząca do wprowadzania do białka nieprawidłowych aminokwasów i w rezultacie do powstania heterogennej puli białek [15]) rozwiązanie to jest powszechnie i z powodzeniem stosowane w wielu laboratoriach [34].

W tym miejscu należy przytoczyć dane eksperymentalne podkreślające istotę problemu rzadko używanych kodonów. Poziom niektórych białek po optymalizacji kodonów genów je kodujących nie tylko nie zwiększa

Tabela V

Porównanie częstości występowania wybranych kodonów u *E. coli*, *C. jejuni* i człowieka. Sporządzono na podstawie [25], <http://www.kazusa.or.jp/codon/>

Kodon (aminokwas)	Częstość występowania (na 1000 kodonów)		
	<i>E. coli</i> K12	<i>C. jejuni</i> 81176	<i>Homo sapiens</i>
AAG (Lys)	10.3	14.4	31.9
GAG (Glu)	17.8	13.3	39.6
CGU (Arg)	21.0	3.9	4.6
CCG (Pro)	23.3	1.6	6.9
GCG (Ala)	33.8	4.0	7.4

się, ale może być niemożliwa [48]. Na poziom ekspresji mogą wpływać także takie czynniki, jak np. lokalizacja rzadko używanych kodonów w genie. Im bliżej 5'-końca genu, tym bardziej rzadkie kodony utrudniają jego ekspresję.

1.3.2. Stabilność mRNA

Średni czas półtrwania mRNA *E. coli* w temperaturze 37°C waha się od kilku sekund do dwudziestu minut, i jest czynnikiem silnie wpływającym na poziom ekspresji genów [34]. Jak do tej pory nie stosuje się czynników zwiększających stabilność mRNA w celu podniesienia poziomu ekspresji obcych genów – wyjątkiem jest stosowanie szczepów niosących mutację w genie kodującym RNazę E. Wyniki niektórych badań sugerują, iż dołączenie stabilizujących sekwencji na 5' i 3' końcu transkryptu może podwyższyć poziom ekspresji, technik tych nie stosuje się jednak standardowo.

Pomocnicze zastosowanie w nadekspresji znalazły rybonukleazy. Jeden z takich systemów pozwala zamienić komórki *E. coli* w swego rodzaju bioreaktory produkujące jedno jedyne białko [14]. Do tego celu wykorzystywane jest białko MazF, będące endorybonukleazą tnącą mRNA w miejscu ACA. Tak więc po indukcji ekspresji genu *mazF* praktycznie cały komórkowy mRNA ulega fragmentacji, całkowicie blokując syntezę białek w komórce. Gdy sekwencja nukleotydowa genu kodującego rekombinowane białko zostanie odpowiednio zmieniona i motywy ACA zostaną z niej usunięte (jest to możliwe bez zmiany sekwencji aminokwasowej, niezależnie od tego, w której ramce odczytu ACA się znajduje), syntetyzowane jest prawie wyłącznie rekombinowane białko. Nadekspresja jest przeprowadzana w niskiej temperaturze (15°C); wzrost bakterii zostaje w tych warunkach całkowicie zahamowany, natomiast pozostają one cały czas aktywne metabolicznie.

1.3.3. Reguła N-końca

Innym czynnikiem wpływającym na stabilność białka jest typ aminokwasu następującego po N-końcowej metioninie [27], gdyż decyduje on o usunięciu lub pozostawieniu N-końcowej formylometioniny. Odcięcie N-formylometioniny jest katalizowane przez formylometionylaminopeptydazę, która odcina ją tym wydajniej, im mniejszy jest łańcuch boczny następującego po niej aminokwasu. N-formylometionina nie jest odcinana, jeżeli występuje po niej histydyna, glutamina, kwas glutaminowy, fenyloalanina, metionina, lizyna, tyrozyna, tryptofan lub arginina.

Aminokwas znajdujący się na N-końcu białka silnie wpływa na jego stabilność. Według reguły N-końca arginina, lizyna, fenyloalanina, leucyna, tryptofan i tyrozyna poprzez stymulację odcięcia N-końcowej formylometioniny destabilizują białko i powodują, że jego czas półtrwania wynosi jedynie dwie minuty [2].

2. Strategie eksperymentalne

2.1. Lokalizacja

Podczas projektowania eksperymentu nadekspresji konkretnego genu niezwykle istotny jest wybór docelowej lokalizacji białka. Każda z lokalizacji ma swoje wady i zalety, które powinny być rozważone przed konstrukcją wektora do ekspresji. Białka nadprodukcowane w *E. coli* mogą być kierowane do cytoplazmy, peryplazmy, bądź wydzielane do podłoża, mogą też wbudowywać się w błonę wewnętrzną (białka błony zewnętrznej są najczęściej oczyszczane z ciał inkluzyjnych [4]).

2.1.1. Cytoplazma

Cytoplazma jest dla większości zastosowań optymalnym przedziałem komórkowym do produkcji rekombinowanych białek ze względu na wysoką wydajność ekspresji [34]. Jednakże ekspresja w tym przedziale komórkowym ma kilka znaczących wad. Jedną z nich jest brak możliwości wprowadzania mostków disiarczkowych; ten problem może zostać rozwiązany poprzez zastosowanie szczepów zdolnych do wprowadzania wiązań między cysteinami w cytoplazmie (szczepy z mutacjami w genach *trxB* i *gor*, jak Novagen Origami). Inną relatywnie często pojawiającą się sytuacją jest agregacja białek w tzw. ciała inkluzyjne [3]. W zależności od przeznaczenia rekombinowanego białka może to być wadą lub zaletą tej lokalizacji.

Ciała inkluzyjne to strukturalnie złożone agregaty niesfałdowanych białek, widoczne w mikroskopie świetlnym [34]. Uzyskanie z nich prawidłowo sfałdowanego białka jest możliwe, aczkolwiek jest to proces pracochłonny i nie zawsze wydajny, a uzyskane w ten sposób białka mogą nie odzyskać aktywności [23]. Prawdopodobieństwo powstawania ciał inkluzyjnych zmniejsza dołączenie do białka hydrofilowego znacznika, koekspresja białek opiekuńczych lub zmiana warunków hodowli, np. obniżenie temperatury, zmiana pH czy składu podłoża (niektóre dodatki, jak sorbitol, rafinoza czy sacharoza) [41].

Wytwarzanie ciał inkluzyjnych ma też kilka zalet. Przede wszystkim ulokowane w nich białka są w znacznym stopniu chronione przed proteolizą [34]. W przypadku syntezy białek toksycznych dla gospodarza może być to jedyny sposób uzyskania nadekspresji, gdyż białka w ciałach inkluzyjnych tracą swoją potencjalną toksyczność [23]. Ciała inkluzyjne są też z powodzeniem stosowane do stymulowania odpowiedzi immunologicznej jako doustna szczepionka [20].

2.1.2. Białka błonowe

Białka błonowe pod względem struktury trzeciorzędowej dzielą się na dwie klasy: zbudowane z wiązki α helis, oraz o strukturze beta-beczulki (β -barrel),

obecne jedynie w błonie zewnętrznej bakterii gramujemnych [45], i stanowiące jedynie niewielki procent wszystkich białek błonowych [4]. Białka o strukturze beta-beczulki są relatywnie łatwe do oczyszczenia, gdyż często, przy zastosowaniu odpowiednich konstrukcji (usunięcie sekwencji sygnałnej) lokują się one w cytoplazmie, tworzą ciała inkluzyjne, z których mogą zostać wyizolowane i ponownie sfałdowane [4]. W przypadku białek o strukturze wiązek helis ponowne fałdowanie po izolacji z ciał inkluzyjnych bardzo rzadko kończy się sukcesem, dlatego też najczęściej podejmowane są próby utrzymania ich lokalizacji w błonie wewnętrznej, a następnie izolacji przy użyciu detergentów [45], co nadal mimo opracowywania nowych systemów jest bardzo trudne. Napotykanym problemem jest bardzo wiele, począwszy od zbyt szybkiego tempa syntezy, a więc i mniejszej wydajności prawidłowego fałdowania, poprzez fakt, iż część białek błony wewnętrznej wymaga obecności konkretnych lipidów w błonie (np. ekspresję występującego u ssaków białka SERT można uzyskać jedynie w obecności cholesterolu, którego nie ma w błonach prokariotycznych), kończąc na aktywacji odpowiedzi na stres w momencie tzw. „przeładowania” błony rekombinowanymi białkami.

Białka błonowe stanowią znaczny procent wszystkich białek komórkowych – u niektórych gatunków nawet do 25% całego proteomu [11]. Oczyszczanie białek błonowych stanowi często poważne wyzwanie. W zależności od planowanych zastosowań rekombinowanego białka można eksprymować jedynie odcinek genu kodujący fragment białka, np. domenę peryplazmatyczną (stosowane w produkcji przeciwciał), i kierować ją do cytoplazmy lub peryplazmy, bądź też próbować wytwarzać białko zlokalizowane w błonie. Jest to rozwiązanie zdecydowanie trudniejsze, ale w niektórych przypadkach jedyne możliwe (np. w przypadku badań struktury).

Stosowane rozwiązania obejmują zmniejszenie tempa translacji poprzez obniżenie temperatury lub/i stężenia substancji indukującej, a także zastosowanie słabszego promotora. Koekspresja genów kodujących białka opiekuńcze może pomóc w fałdowaniu i zapobiec akumulacji nieprawidłowo zwiniętych form, jednak optymalizacja warunków koekspresji musi być empiryczna, jest procesem żmudnym i długotrwałym. Kolejną możliwością jest zastosowanie szczepów *E. coli*, które ze względu na bliżej nieokreślone mutacje (prawdopodobnie wpływające na poziom polimerazy RNA faga T7 [11]) są mniej wrażliwe na obecność toksycznych białek i okazały się pomocne w nadprodukcji białek błonowych (szczepy C41 i C43).

Jeżeli wprowadzenie drobnych zmian do oczyszczanego białka nie uniemożliwia jego późniejszego zastosowania, pewne modyfikacje jego sekwencji ami-

nokwasowych mogą zwiększyć szanse na wydajną nadprodukcję. Często pojawiają się trudności z translokacją N-końca białka do peryplazmy. Łatwość i wydajność translokacji zależy od tego, czy N-koniec może pozostać niesfałdowany (jest to tym łatwiejsze, im jest on krótszy), a także od liczby dodatnio naładowanych aminokwasów i tzw. „siły” pierwszego segmentu transbłonowego (tzn. długości, wypadkowego ładunku oraz hydrofobowości) [45]. Wiele fragmentów N-końcowych wykazuje preferencje do pozostawiania w cytoplazmie *E. coli*, tym samym utrudniając, a nawet uniemożliwiając wbudowanie się białka w błonę wewnętrzną. Rozwiązaniem może być skonstruowanie fuzji N-końca białka z sekwencją sygnałną bądź innym białkiem, jak na przykład białkiem wiążącym maltozę (MBP). Rozwiązaniem problemu może okazać się zidentyfikowane stosunkowo niedawno białko Mistic z *B. subtilis*, które, dołączone jako białko fuzyjne na N-końcu, efektywnie ułatwia wbudowanie się fuzji w błonę [31].

2.1.3. Peryplazma

Białka przeznaczone do transportu poza cytoplazmę są syntetyzowane jako preproteiny zawierające N-końcową sekwencję sygnałną, która zostaje odcięta podczas translokacji przez błonę wewnętrzną, w ich translokację zaangażowany jest translokon Sec (secretion) lub Tat (twin arginin translocation) [3, 18]. Typowa sekwencja sygnałna białek *E. coli* składa się z 18–30 aminokwasów, w tym kilku zasadowych na N-końcu, hydrofobowego rdzenia, oraz C-końcowego motywu rozpoznawanego przez sygnałowe proteazy. Dołączenie takiej sekwencji sygnałnej do rekombinowanego białka umożliwia jego transport poza cytoplazmę. Najczęściej stosowane sekwencje sygnałne pochodzą z białek OmpT, OmpA, PelB, PhoA, MalE, LamB czy β -laktamazy [34]. Dostępne są także systemy jednocześnie umożliwiające transport przez błonę wewnętrzną, jak i zwiększoną wydajność wprowadzania mostków disiarczkowych w peryplazmie. W takich konstrukcjach oczyszczane białko tworzy fuzję z oksydoreduktazą tiolowo-disiarczkową, DsbA lub DsbC.

Istotną zaletą kierowania białek do peryplazmy jest niższy stopień zanieczyszczenia preparatu innymi białkami, co ułatwia procedury oczyszczania [34]. Prawdopodobieństwo proteoliznej degradacji jest niższe, a uzyskania natywnej konformacji N-końca białka – wyższe. Ponieważ peryplazma jest środowiskiem utleniającym, możliwe jest tam wprowadzanie mostków disiarczkowych, a więc zwiększa się możliwość otrzymania białka o natywnej konformacji [3]. System transportu do peryplazmy nie zawsze musi być wydajny, czasem występuje tzw. „zapychanie błon”, istnieje także ryzyko tworzenia się ciał inkluzyjnych w przestrzeni peryplazmatycznej [16].

2.1.4. Sekrecja do podłoża

Zaletą sekrecji rekombinowanych białek do podłoża jest niewielki stopień zanieczyszczenia innymi białkami [16]. Ułatwia to znacznie oczyszczanie rekombinowanego białka. Laboratoryjne szczepy *E. coli* nie posiadają efektywnych mechanizmów transportu przez błonę zewnętrzną [34]. W celu uzyskania rekombinowanego białka w podłożu, stymuluje się pasywny transport przez błonę zewnętrzną. Można to osiągnąć poprzez destabilizację elementów strukturalnych osłon komórkowych *E. coli*, np. stosując szczepy nie posiadające niektórych elementów struktury błon. Inna strategia polega na wprowadzeniu do laboratoryjnego szczepu *E. coli* genów kodujących białka budujące aparaty sekrecji pochodzących z patogennych szczepów *E. coli* np. genów kodujących białka budujące aparat sekrecji typu I (eksport hemolizyny). Białko jest transportowane z cytoplazmy bezpośrednio na zewnątrz komórki poprzez kanał białkowy przechodzący przez błonę wewnętrzną i zewnętrzną [8]. Kierowanie rekombinowanych białek do peryplazmy może w określonych warunkach prowadzić do ich niekontrolowanego wydzielenia do podłoża; takie podejście eksperymentalne pozwoliło na oczyszczenie podjednostki B toksyny cholery [33]. Ciekawym pomysłem jest zastosowanie tzw. form L, czyli szczepów pozbawionych błony zewnętrznej oraz otoczki mureinowej [9]. W tym systemie białka dołączone do standardowej sekwencji sygnałnej zostają, poprzez błonę wewnętrzną, wydzielone bezpośrednio do podłoża. Jednakże, mimo pewnych osiągnięć w tej dziedzinie, oczyszczanie białek wydzielonych do podłoża nie jest procedurą stosowaną standardowo [34].

2.2. Strategie klonowania

Pewnych trudności – nawet znanych z teorii – nie można przewidzieć podczas projektowania konstruktów do ekspresji. Na przykład oczyszczane białko bądź jego fragment może nie ulegać ekspresji, tworzyć ciała inkluzyjne, okazać się trudne do oczyszczenia, bądź też nie wykazywać pożądanej aktywności. Dlatego właśnie tak ważne są metody szybkiego i efektywnego tworzenia wielu konstruktów, umożliwiające wypróbowanie kilku wariantów klonowania i znalezienie optymalnego dla konkretnego białka.

Dany gen bądź jego fragment jest zazwyczaj najpierw klonowany do standardowego wektora nieekspresyjnego (jak pBluescript czy ułatwiający klonowanie produktów PCR pGEM), a następnie sekwencjonowany. Taka strategia ma kilka zalet; jeżeli produkt klonowanego genu jest toksyczny dla komórki, jego wklonowanie prosto do wektora ekspresyjnego może okazać się trudne lub nawet niemożliwe. Wektory ekspresyjne są często niskokopijne, co utrudnia klonowanie i se-

kwencjonowanie badanego genu. Ważnym argumentem przemawiającym na korzyść takiego podejścia jest także możliwość sprawdzenia poprawności sekwencji nukleotydowej klonowanego genu. Dopiero po sprawdzeniu sekwencji nukleotydowej, gen może być przeklonowany do wybranego wektora ekspresyjnego.

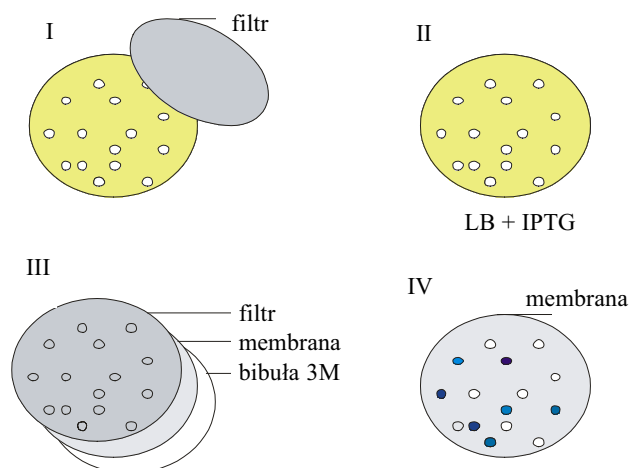
Większość konstruktów do nadekspresji białek jest tworzona za pomocą enzymów restrykcyjnych i liganzy (tzw. metoda REaL) [17]. Jest to metoda najmniej kosztowna, a większość szeroko stosowanych wektorów ekspresyjnych jest do niej dostosowana. Problemy mogą pojawiać się, gdy w klonowanym genie występują miejsca restrykcyjne popularnie stosowane w miejscach do klonowania wektorów. Jest to także metoda relatywnie pracochłonna.

Alternatywą dla metody REaL jest technika wykorzystująca rekombinację specyficzną co do miejsca. Stosowane w niej enzymy przeprowadzają cięcie i ligację fragmentu DNA podczas jednego etapu, dzięki czemu reakcja jest bardzo wydajna. Za wysoką specyficzność przeprowadzanej reakcji odpowiada fakt, iż miejsca zajścia rekombinacji są dużo dłuższe niż miejsca rozpoznawane przez enzymy restrykcyjne. Dzięki wysokiej wydajności i specyficznosci metoda ta może być wykorzystywana do klonowania bardzo długich fragmentów DNA. Obie dostępne komercyjnie metody wykorzystujące zjawisko rekombinacji specyficznej co do miejsca, Gateway (Invitrogen) i Creator (ClonTech), są oparte na systemach fagowych, pochodzących odpowiednio z faga lambda i faga P1. Wadą tych technik są jednak wysokie koszty.

2.2.1. Projektowanie konstruktów

Ze względu na fakt, iż trudno jest przewidzieć, jaki konstrukt będzie optymalny w procesie oczyszczania konkretnego białka, skuteczną strategią może być jednoczesne stworzenie kilku konstruktów i eksperymentalne dobranie optymalnych warunków nadekspresji genu i oczyszczania jego produktu [28]. Wiele wektorów ma kompatybilne miejsca restrykcyjne i jednoczesne klonowanie nie wymaga oddzielnego przygotowywania klonowanych fragmentów DNA. Zastosowanie metody Gateway również umożliwi łatwe klonowanie do wielu wektorów jednocześnie.

W pracy przeglądowej P e t i i P a g e [28] przedstawili przykładową strategię eksperymentalną prowadzącą do oczyszczenia konkretnego białka. Gen kodujący to białko jest klonowany w trzech różnych wektorach; przedstawione na schemacie wektory zawierają znacznik Thio₆ (fragment DNA kodujący pierwsze sześć aminokwasów tioredoksyny, pETRP1), gen kodujący S-transferazę glutationu (pETM-30), oraz gen kodujący białko wiążące maltozę (pETM-41). W większości przypadków, co najmniej jeden z tych znaczników powinien umożliwić nadprodukcję białka



Rys. 2. Filtracja kolonijna

Komórki zawierające różne warianty konstruktów do ekspresji są wysiewane na podłożu LB (I). Na powierzchni kolonii umieszczany jest filtr membranowy. Kolonie przylegają do niego i mogą być oderwane od podłoża. Ekspresja białka jest indukowana poprzez umieszczenie filtru na szalce z podłożem LB zawierającym IPTG (II). Następnie komórki są poddawane lizie, a białka rozpuszczalne dyfundują poprzez filtr do membrany nitrocelulozowej, podczas gdy ciała inkluzyjne pozostają na filtrze (III). Kolonie ekspresyjnie rozpuszczalne białko mogą być zidentyfikowane na membranie nitrocelulozowej za pomocą standardowych metod immunologicznych (IV).

w formie rozpuszczalnej. Na N-końcu rekombinowanego białka dołączony jest jeszcze dodatkowo znacznik His₆, umożliwiający oczyszczanie białka w standardowych warunkach metodą chromatografii powinowactwa na jednym typie złoża.

Jeżeli pierwsza pula konstruktów nie da oczekiwanych rezultatów, rozwiązaniem może być zastosowanie innych, rzadziej używanych znaczników. Może być to Z-tag (tzw. domena Z z *Staphylococcus* sp., pET22-1a), GB1-Tag (domena β1 białka G z *Streptococcus* sp., pHISGB1) czy białko NusA.

Bardzo istotnym elementem eksperymentów nadprodukcji białek jest opracowanie metodyki przeszukiwania (tzw. screening), czyli szybkiego i nie wymagającego dużych nakładów czasu na sprawdzenie, który z konstruktów jest optymalny do oczyszczania białka na dużą skalę. Jednym z najczęściej stosowanych sposobów jest elektroforeza w żelu poliakrylamidowym (SDS-PAGE). Na żelu rozdziela się najczęściej białka pochodzące z rozpuszczalnej frakcji lizatu komórek z małej ilości hodowli; stosowane jest też wstępne oczyszczenie na małej ilości złoża.

Metodę tą można, ze względu na koszty i czas, zastosować tylko do pewnej liczby prób [10]. Alternatywną metodą jest tzw. filtracja kolonijna (colony filtration, Co-Fi). Ta technika nie wymaga przygotowywania lizatów z hodowli, obecność rozpuszczalnego białka można stwierdzić już na poziomie kolonii, co umożliwia przeszukanie znacząco większej liczby konstruktów. Szczegóły metody są pokazane na rys. 2.

3. Podsumowanie

Od końca poprzedniego wieku wyraźnie zauważamy zmianę strategii analizy procesów genetycznych. Punkt ciężkości przesuwają się z badania pojedynczych genów czy operonów w kierunku globalnych analiz komórek zarówno pro- jak i eukariotycznych. Dodatkowo w ostatnich latach coraz więcej uwagi poświęcamy nie analizom genów czy transkryptów (genomika, transkryptomika) a analizom białek i ich oddziaływań (proteomika, interaktomika). Tylko niewiele białek występują w komórkach naturalnych gospodarzy w dostatecznie dużej ilości aby móc je oczyszczać do badań funkcjonalnych lub/i strukturalnych. W ostatnich latach opracowano wiele nowych strategii pozwalających na

Tabela VI

Najczęściej napotykane problemy i ich możliwe rozwiązania w eksperymentach nadekspresji białek w komórkach *E. coli*.
Sporządzono na podstawie [23, 30, 41]

Problem	Możliwa przyczyna	Rozwiązania
Śmierć komórek, brak kolonii	Białko toksyczne, zbyt wysoki poziom ekspresji	Silniejsza kontrola ekspresji podstawowej Słabszy promotor. Niższa temperatura Mniejsze stężenie induktora
Białko w ciałach inkluzyjnych	Zbyt wysoki poziom ekspresji	Słabszy promotor. Niższa temperatura Mniejsze stężenie induktora Plazmid o niższej liczbie kopii Dołączenie hydrofilowego znacznika
Brak ekspresji, ekspresowane tylko fragmenty białka	Problem używalności kodonów	Zastosowanie gospodarza dostarczającego rzadkie tRNA. Niższa temperatura
Białko nie jest transportowane do peryplazmy	Sekwencja sygnałowa nie rozpoznawana w <i>E. coli</i> Zbyt mało wydajne mechanizmy translokacji	Dołączenie innej sekwencji sygnałowej bądź znacznika. Koekspresja genów <i>sec</i>
Brak ekspresji	Proteoliza (szczególnie w przypadku białek mniejszych niż 10 kDa) Mutacje wprowadzające kodon stop wewnątrz sekwencji kodującej genu	Niższa temperatura. Krótszy czas indukcji Dołączenie znacznika. Sekwencjonowanie

zwiększenie wydajności procesu wytwarzania i oczyszczania białek. Komórki kilku gatunków bakterii przystosowano, na drodze manipulacji genetycznych, do wydajnej produkcji heterologicznych białek. Największy postęp niewątpliwie osiągnięto w adaptacji do tego celu komórek *E. coli*. Tym niemniej jeszcze często, szczególnie w wypadku białek błonowych, stanowiących około 30% białek komórki stawiany cel jest trudny do osiągnięcia. Tabela VI przedstawia najczęściej napotymane trudności przy nadprodukcji rekombinowanych białek w komórkach *E. coli* i potencjalne możliwości ich przewyciężenia. Otrzymanie dostatecznej ilości białka zachowującego właściwą strukturę jest często warunkiem rozpoczęcia badań o charakterze zarówno poznawczym jak i aplikacyjnym białek. Opracowywanie nowych leków i preparatów immunoprolaktycznych (strukturalna wakcynologia) coraz częściej bazuje na rozwiązanej strukturze białka. Tabela VI przedstawia najczęściej napotymane trudności przy nadprodukcji rekombinowanych białek w komórkach *E. coli* i potencjalne możliwości ich przewyciężenia.

Piśmiennictwo

1. Arnau J., Lauritzen C., Petersen G.E., Pedersen J.: Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* **48**, 1–13 (2006)
2. Bachmair A., Finley D., Varshavsky A.: In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186 (1986)
3. Baneyx F.: Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411–421 (1999)
4. Bannwarth M., Schulz G.E.: The expression of outer membrane proteins for crystallization. *Biochim. Biophys. Acta*, **1610**, 37–45 (2003)
5. Butt T.R., Edavettal S.C., Hall J.P., Mattern M.R.: SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **43**, 1–9 (2005)
6. Chart H., Smith H.R., La Ragione R.M., Woodward M.J.: An investigation into the pathogenic properties of *Escherichia coli* strains BLR, BL21, DH5 α and EQ1. *J. Appl. Microbiol.* **89**, 1048–1058 (2000)
7. Chayen N.E.: Turning protein crystallization from an art into a science. *Curr. Opin. Struct. Biol.* **14**, 577–583 (2004)
8. Choi J.H., Lee S.Y.: Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **64**, 625–635 (2004)
9. Cornelis P.: Expressing genes in different *Escherichia coli* compartments. *Curr. Opin. Biotechnol.* **11**, 450–454 (2000)
10. Cornvik T., Dahlroth S.L., Magnusdottir A., Herman M.D., Knaust R., Ekberg M., Nordlund P.: Colony filtration blot: a new screening method for soluble protein expression in *Escherichia coli*. *Nature Methods*, **2**, 507–509 (2005)
11. Cunningham F., Deber C.M.: Optimizing synthesis and expression of transmembrane proteins. *Methods*, **41**, 370–380 (2007)
12. Dyck M.K., Lacroix D., Pothier F., Sirard M.A.: Making recombinant proteins in animals – different systems, different applications. *Trends Biotechnol.* **21**, 394–399 (2003)
13. Esposito D., Chatterjee D.K.: Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17**, 353–358 (2006)
14. Falzon L., Suzuki M., Inouye M.: Finding one of a kind: advances in single-protein production. *Curr. Opin. Biotechnol.* **17**, 347–352 (2006)
15. Gustafsson C., Govindarajan S., Minshull J.: Codon bias and heterologous protein expression. *Trends Biotechnol.* **22**, 346–353 (2004)
16. Hannig G., Makrides S.C.: Strategies for optimizing heterologous gene expression in *Escherichia coli*. *Trends Biotechnol.* **16**, 54–60 (1998)
17. Hartley J.L.: Cloning technologies for protein expression and purification. *Curr. Opin. Biotechnol.* **17**, 359–366 (2006)
18. Holland I. B.: Translocation of bacterial proteins – an overview. *Biochim. Biophys. Acta*, **1694**, 5–16 (2004)
19. Jenny R.J., Mann K.G., Lundblad R.L.: A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif.* **31**, 1–11 (2003)
20. Kęsik M., Sączyńska V., Szewczyk B., Plucienniczak A.: Inclusion bodies from recombinant bacteria as a novel system for delivery of vaccine antigen by the oral route. *Immunol. Lett.* **91**, 197–204 (2004)
21. Kunji E.R., Slotboom D.J., Poolman B.: *Lactococcus lactis* as host for overproduction of functional membrane proteins. *Biochim. Biophys. Acta*, **1610**, 97–108 (2003)
22. Kunji E.R., Chan K.W., Slotboom D.J., Floyd S., O'Connor R., Monné M.: Eukaryotic membrane production in *Lactococcus lactis*. *Curr. Opin. Biotechnol.* **16**, 546–551 (2005)
23. Makrides S.C.: Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**, 512–538 (1996)
24. Marblestone J.G., Edavettal S.C., Lim Y., Lim P., Zuo X., Butt T.R.: Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* **15**, 182–189 (2006)
25. Nakamura Y., Gojobori T., Ikemura T.: Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292 (2000)
26. Nevalainen K.M., Te'o V.S., Bergquist P.L.: Heterologous protein expression in filamentous fungi. *Trends Biotechnol.* **23**, 468–474 (2005)
27. Novagen. pET system manual: 11th edition, EMD Biosciences, Darmstadt, 2006.
28. Peti W., Page R.: Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein Expr. Purif.* **51**, 1–10 (2007)
29. Poole E.S., Brown C.M., Tate W.P.: The identity of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli*. *EMBO J.* **14**, 151–158 (1995)
30. Qiagen. The QIAexpressionist, Qiagen, 2003
31. Roosild T.P., Greenwald J., Vega M., Castronovo S., Riek R., Choe S.: NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science*, **307**, 1317–1321 (2005)
32. Schumann W., Ferreira L.C.: Production of recombinant proteins in *Escherichia coli*. *Genet. Mol. Biol.* **27**, 442–453 (2004)
33. Slos P., Speck D., Accart N., Kolbe H.V., Schubnel D., Bouchon B., Bischoff R., Kieny M.P.: Recombinant cholera toxin B subunit in *Escherichia coli*: high-level secretion, purification, and characterization. *Protein Expr. Purif.* **5**, 518–526 (1994)

34. Sørensen H.P., Mortensen K.K.: Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115**, 113–128 (2005)
35. Srivastava P., Deb J.K.: Gene expression systems in corynebacteria. *Protein Expr. Purif.* **40**, 221–229 (2005)
36. Stenström C.M., Jin H., Major L.L., Tate W.P., Isaksson L.A.: Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene*, **263**, 273–284 (2001)
37. Stratagene. BL21-CodonPlus competent cells: instruction manual. La Jolla, CA, 2005
38. Studier F.W., Moffatt B.A.: Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986)
39. Studier F.W.: Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005)
40. Terpe K.: Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**, 523–533 (2003)
41. Terpe K.: Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–222 (2006)
42. Terpe K.: Protein-Affinität-Tags. *Biospektrum*, **4**, 389–391 (2007)
43. Villalobos A., Ness J.E., Gustafsson C., Minshull J., Govindarajan S.: Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics*, **7**, 285–292 (2006)
44. Wacker M., Linton D., Hitchen P.G., Nita-Lazar M., Haslam S.M., North S.J., Panico M., Morris H.R., Dell A., Wren B.W., Aebi, M.: N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*, **289**, 1790–1793 (2002)
45. Wagner S., Bader M.L., Drew D., de Gier J.: Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24**, 364–371 (2006)
46. Waugh D.S.: Making the most of affinity tags. *Trends Biotechnol.* **23**, 316–320 (2005)
47. Widmann M., Christen P.: Comparison of folding rates of homologous prokaryotic and eukaryotic proteins. *J. Biol. Chem.* **275**, 18619–18622 (2000)
48. Wu X., Jörnvall H., Berndt K.D., Oppermann U.: Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochem Biophys. Res. Commun.* **313**, 89–96 (2004)